

周报总结

2015-07-12

前周回顾:

上周在做的时候,发现我们的数据分布存在问题,绝大多数的数据其实集中在比较小的一部分范围内(99.9%的数据聚集在 40%左右的空间)。这些点的数目很少,然而因为这些点的存在,导致了我们的数据归一化的时候绝大部分的数据其实是距离很近,因此直接对整个聚类造成了影响。

本周总结:

本周工作主要分为两部分,一是去重新整理数据,重新计算数据,二是本科生陆陆续续来实验室了,开始计划给他们安排一些任务。

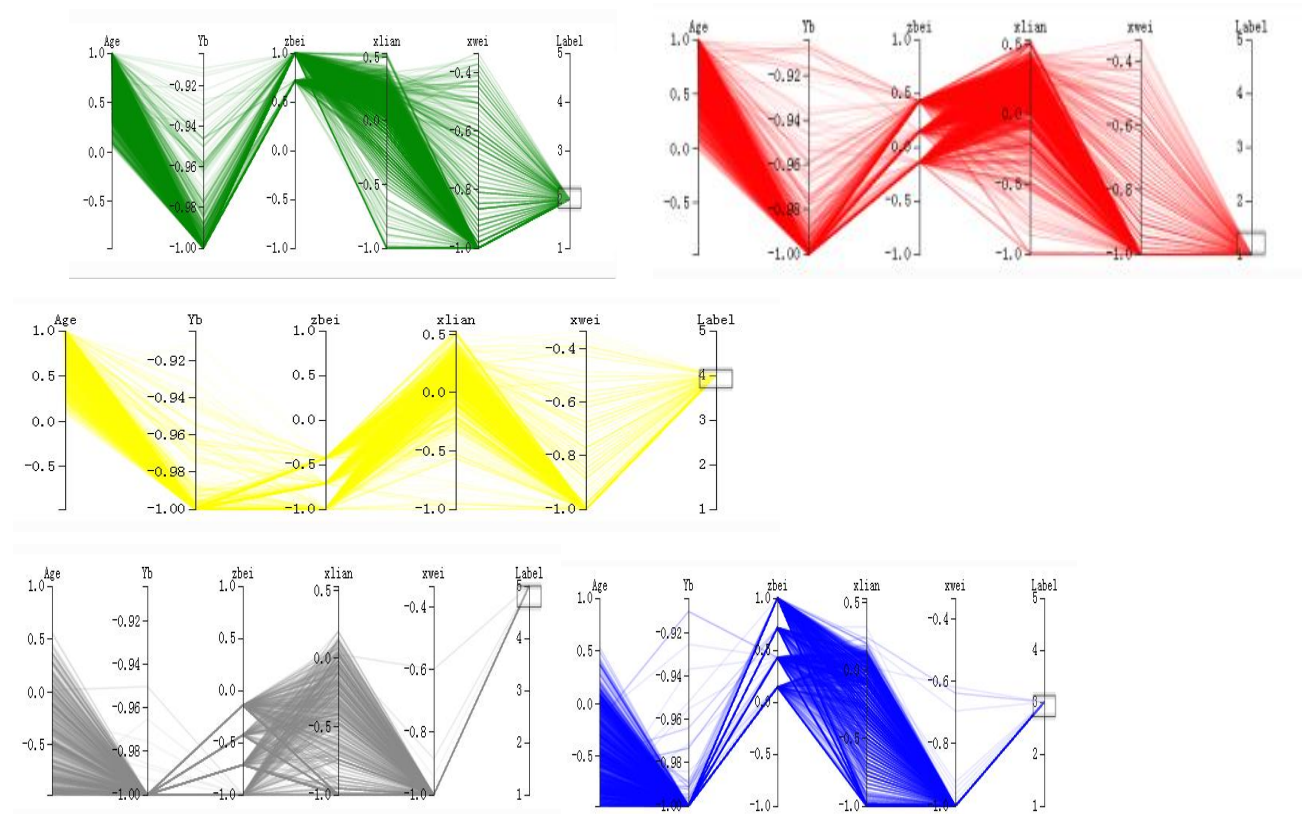
上周做的时候发现我们的数据存在问题,很多数据存在孤立点的情况,本周开始的时候,我把我们用的数据每一个都重新看重新整理了一下,发现不同数据之间其实差别还是有的,有的数据比较集中有的数据比较分散。因此在整理的时候,需要一个比较统一而且有意义的清洗方式。使用的方式是: 1 首先求每一个类的概率密度, 2 根据每一个类的概率密度,我们设定 2 个阈值,一个最大阈值一个最小阈值,分别控制保留数据的最小值和最大值。发现结果有较好的效果。

本周陆陆续续本科生都来实验室了,实验室的又处在了之前投稿时,人满为患的状态。对于本科生,本周都找他们每个人去聊了一下,每个人讲了一下我们的项目,还有接下来可能要做什么。在开始的时候还是以培训为主,之前秘书有发过一个巫老师给新人的培训计划,感觉不错,开始的时候先用这个让他们熟悉一下可视化相关的工作,还有 WEB 端的工作。因为之后的工作很可能在 WEB 端进行。

同时,看了巫老师那个培训计划,深感自己里面还有很多要学的。所以不但给本科生有所安排任务,自己也计划在这段时间把这些文章都看完,同时因为之后可能要用 D3 之类的工具,自己 web 端又不是很了解,所以打算学一些 web 端东西。所以这周开始,给自己增加了学习计划,计划在这段时间,到开学每周都要去学习一些文档里提到的相关内容。还有 JS,WEB 等等。

具体工作

工作这里主要选取了 5 个属性，分别为：玩家游戏年龄，元宝数量、玩家装备等级、玩家修炼等级、玩家修为等级



可以看到聚类比较清晰的将玩家分成了几种不同的类别，上面三个类别为玩家年龄较高的类，下面两个类别为玩家年龄较低的类别。

其中第一个为装备等级较高的类别，第二个为装备等级一般的类别第三个为装备等级较低的类别。同时，装备较好，其修炼的值较装备中等修炼值来说略低(从平行坐标走势中就能看出)。

第四个、第五个类别为玩家在线时长较短的类别，他们的元宝数量普遍不高，第四个普遍来说是一般的低端玩家，因为其装备 修为、修炼值普遍偏低。而第五个装备值较高，然而修炼和修为值不高，由此可见，修炼和修为很可能需要时间去积累才能有较高的水平。

下周安排：

- 1 找巫老师讨论下一步工作。
- 2 学习 JS，阅读文档 week1 部分。
- 3 给本科生开会，布置每周任务。